THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 53, No. 5, May–June 1974
Printed in U.S.A.

Mathematical Analysis of an Adaptive Quantizer

By DEBASIS MITRA

(Manuscript received December 4, 1973)

This paper presents a mathematical analysis of an adaptive quantizer, a pulse code modulator, which is used for coding speech and other continuous signals with a large dynamic range into digital form. The device is a two-bit quantizer in which the step size is modified at every sampling instant with the object of adapting the range of the device to the intensity level of the signal. In the adaptation algorithm analyzed in the paper, the encoded information of the previous sampling instant is used either to increase or to decrease the step size by fixed, but not necessarily equal, proportions.

Initially, the stochastic stability of the device is established by constructing a stochastic Liapunov function. Various basic identities and bounds on aspects of the behavior of the device are obtained. The qualitative results obtained indicate the nature of the trade-offs between the quality of the steady state and the transient performance of the device. Also, formulas are developed for the purpose of evaluating the mean time required for the step size to adapt from arbitrary initial conditions to certain optimal values.

I. INTRODUCTION

A mathematical analysis of an adaptive quantizer is presented in this paper. The coding thresholds of the device, also referred to as the step sizes, are not fixed but adapt according to a particular alogrithm. The object of the algorithm is to modify the threshold to larger or smaller levels, depending on whether the signal intensity level is high or low, in a manner that allows a decoder at the receiving end to effectively reconstruct the continuous signal. The basic two-bit quantizer, i.e., quantizers with four output levels with codes 01, 00, 10, and 11, is characterized by a particular function of the following form at each sampling instant.

Input refers to the nth sample of the continuous signal, $x(n)$, $n = 0, 1, 2, \cdots$; output refers to the coded signal to be transmitted at that instant; and Δ is the step size. In adaptive quantizers of the type to be investigated here, the step size is variable and the step size at the nth sampling instant is denoted by $\Delta(n)$. The step size uniquely defines the entire function in the manner indicated by Fig. 1; hence, the complete adaptive quantizer is associated with a sequence of functions. The adaptive quantizers that are the subject of this paper are basically characterized by the following adaptation algorithm

$$\Delta(n+1) = M_1\Delta(n) \quad \text{if} \quad |x(n)| \leqq \Delta(n) \tag{1a}$$

$$= M_2\Delta(n) \quad \text{if} \quad |x(n)| > \Delta(n), \tag{1b}$$

where M_1 and M_2, called multiplier coefficients, are fixed constants satisfying* $0 < M_1 < 1 < M_2$. Variations on (1) are considered in the main text, although the discussion in the introductory section is in terms of (1). Results on adaptive quantizers with output levels more numerous than 4 will be considered in a future publication.

The adaptation algorithm in (1) is due to Cummiskey, Flanagan, and Jayant.[1,2] In Ref. 1 Jayant presents the results of extensive computer simulations undertaken to determine the multiplier coefficients which maximize various performance functionals. A class of random inputs $\{x(n)\}$ that is considered is obtained by passing a discrete, white, Gaussian process through a filter with a single pole. In Ref. 2, Cummiskey, Jayant, and Flanagan consider a differential PCM coder in which the adaptive quantizer is used together with a fixed first-order predictor in the feedback loop. Their work has its direct antecedents in the various schemes[3,4,5] for adapting step sizes in delta-modulators, a one-bit quantizer, and in the work of Wilkinson.[6] Wilkinson's paper on a two-bit adaptive quantizer, largely concerned with hardware implementation, is particularly interesting. In his scheme, the step size is controlled by a moving fraction obtained by keeping a tally of the number of times the input falls in the lower slot of the quantizer. Goodman and Gersho[7] have independently looked at the adaptive quantizer from a theoretical standpoint and their work complements the work described here.

In this paper we make a number of simplifying assumptions about the input sequence $\{x(n)\}$, the most restrictive being the assumption

---

*Since the absolute value of the input in Fig. 1 is partitioned into $[0, \Delta]$ and $(\Delta, \infty]$, we shall loosely refer to the event leading to (1a) as "the input falling in the lower slot."

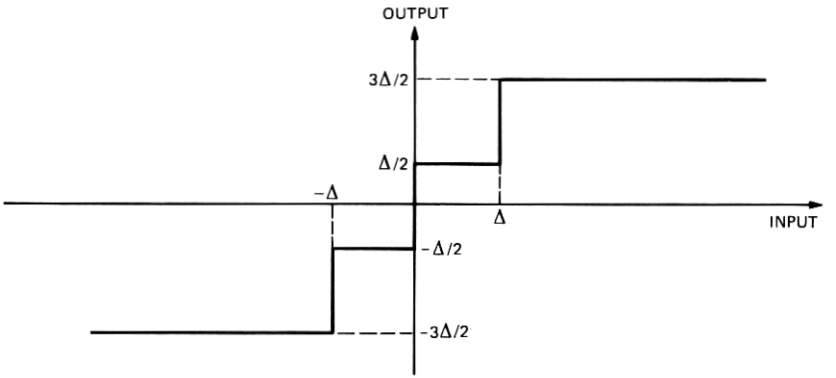

Fig. 1—The quantizer function.

that it is a sequence of independent random variables. However, we have obtained for the idealized model precise results which indicate rather fully the trade-offs involved in the choice of the multiplier coefficients. Also, we have developed formulas for efficiently computing functionals as aids in the design problem. We believe that the broad qualitative features of the device that are found to hold in this model carry over for more realistic input processes. It is hoped too that the techniques developed here will provide a point of reference for future work.

The mathematical analysis, for the main part, is of a random walk on the integers, whose complexity is due to the dependence of the state transition probabilities on the states. The structure of the random walk which is exploited here is rather general, and for this reason the model is of independent interest; to our knowledge, the main mathematical results have not appeared in the literature on random walks.

The organization of the paper is as follows. In Section 1.1 we continue the discussion on the adaptation algorithm in the context of a particular idealized model of the sequence $\{x(n)\}$, and we discuss some of the results to be derived later and what is already known about optimal quantization in the nonadaptive framework. In Sections 1.2 and 1.3 we give the basic equations of the process arising from (1), and certain modifications of it, when the input sequence $\{x(n)\}$ is independent and identically distributed. In Section II the stochastic stability of the device is established under general conditions. The existence and uniqueness of the stationary distribution of the step size is proved by constructing a stochastic Liapunov function for the random process. Section III examines in detail the stationary step

size distribution. In Section 3.2 we prove an identity which explicitly gives the stationary probability of the input falling in the lower slot of the quantizer, i.e., $\Pr_s [|x(n)| \leqq \Delta(n)]$. In Section 3.3 sharp bounds are obtained on the stationary probabilities. It is shown that for almost all values of the multiplier coefficients there exists a natural center of the distribution and that the stationary probabilities fall off at least geometrically with increasing distances from the natural center. In Section 3.5 results are obtained on a particular limiting behavior, namely, the effect of the stationary distribution of making both multiplier coefficients close to unity. Section IV is devoted to the transient response of the device. In Section 4.1 we develop formulas for the efficient computation of the time required for the step size to adapt from an arbitrary initial value to the desired step size. Section 4.2 by giving an explicit bound on this time provides some insight into the dependence of the adaptive time on the choice of the multiplier coefficients. Finally, we report some computational results.

1.1 Background

In an idealized model for the samples, $x(n)$, of the continuous signal process, assume that $\{x(n)\}$ is a sequence of independent random variables with zero mean. Assume further that the distribution of $x(n)$ for every n is an element of the same equivalence class of distributions in which the distributions are equivalent to within a scaling operation. The scaling or intensity level changes slowly with n. For instance, the equivalence class of distributions may be the family of Gaussian distributions and only the variance, indicating the intensity level, changes with n.

It is necessary to recall at this stage some known facts concerning the design of quantizers in the nonadaptive framework[8] where $\{x(n)\}$ is a sequence of independent, identically distributed random variables and the step size is fixed. Suppose that $E[\{y(n) - x(n)\}^2]$ measures the performance of the quantizer where $y(n)$ is the nth output of the device.* The step size which minimizes this functional, $\hat{\Delta}$, is in principle easy to establish, and $\hat{\Delta}$ is uniquely characterized by the probability of the input falling in the lower slot, i.e., $\Pr [|x(n)| \leqq \hat{\Delta}]$. Another observation that is equally easy to verify is that the optimal step size has the property that if the distribution of $\{x(n)\}$ is scaled, then the optimal step size is obtained by an identical scaling of the previous optimal step size. A convenient way of stating this observation is: a

* It is not essential that the performance functional be of that form.

property of the optimal step size that is invariant to scaling of the distribution of $\{x(n)\}$ is the probability that the absolute value of the input $x(n)$ does not exceed the optimal step size. For instance, when the distribution is Gaussian it is known that this probability is close to 0.68.[8]

An intermediate step in proceeding from the nonadaptive case to the more general model described prior to it, in which the identically distributed condition does not hold, is provided by the following model. Assume that the sequence $\{x(n)\}$ is indeed independent and identically distributed, and that the equivalence class of distributions to which the particular distribution belongs is known. However, the scaling parameter is unknown. It is relatively straightforward to state the requirements on a well-behaved algorithm operating in this simple framework, and, if these requirements are always satisfied, then it is possible to conclude that the device will operate satisfactorily for the more general model. The requirements are: (*i*) for arbitrary initial step size guesses, the step size rapidly converges to the optimal step size, and (*ii*) it is thereafter localized in a small neighborhood of that point. This paper separately analyzes the two requirements in the simple framework just described. Considerations related to (*i*) and (*ii*) are lumped respectively under the terms "transient response" and "steady-state response," since the latter property is effectively investigated in terms of the stationary distribution of the step size, assuming one exists. A good reason for the division is that they lead, in some ways, to quite opposite requirements for the multiplier coefficients.

Consider, in the light of what is known about optimal quantization in the nonadaptive framework, what is required for the localization property, requirement (*ii*), to hold. When the stationary distribution has both of the following properties, it is possible to establish an effective correspondence and infer that (*ii*) holds: (*a*) the stationary probability of the step size falling in the lower slot, i.e., $\Pr_s [|x(n)| \leqq \Delta]$ equals the known value associated with the particular family of distributions; and (*b*) the mass of the stationary distribution is concentrated in the small neighborhood of a point. In Section III we show that by appropriate choice of the multiplier coefficients it is possible to achieve both requirements.

1.2 Basic assumptions and equations

We consider only quantizers with multiplier coefficients having the following structure:

$$M_1 = \gamma^{-k} \quad \text{and} \quad M_2 = \gamma^{l}, \tag{2}$$

where γ is some real number greater than 1 and k and l are positive integers. We shall further make k and l relatively prime, i.e., their greatest common factor is 1. If, as we shall assume, the initial step size is of the form γ^i, with i an integer, then the step size is always of that form and the space of possible step sizes forms a lattice.*

There is a step size with, as we shall see, certain claims to being the central step size for a particular distribution of $\{x(n)\}$ and choice of parameters k and l; this step size is used as a reference point. There exists an integer $\bar{i}$ such that†

$$\Pr[|x(n)| \leqq \gamma^{\bar{i}-1}] < \frac{l}{k+l} \leqq \Pr[|x(n)| \leqq \gamma^{\bar{i}}]. \tag{3}$$

We denote $\gamma^{\bar{i}}$ by C and refer to it as the *central step size*; all step sizes are considered to be of the form $C\gamma^i$, $i = 0, \pm 1, \pm 2, \cdots$.

Obviously, it is more convenient to work with the log transform of the step size, so let

$$\omega(n) \triangleq \log_\gamma \Delta(n) - \log_\gamma C. \tag{4}$$

From the original algorithm we have

$$\begin{aligned} \omega(n+1) &= \omega(n) - k \quad \text{if} \quad |x(n)| \leqq C\gamma^{\omega(n)} \\ &= \omega(n) + l \quad \text{if} \quad |x(n)| > C\gamma^{\omega(n)}. \end{aligned} \tag{5}$$

We have in (5) a Markov chain with states $0, \pm 1, \pm 2, \cdots$. The state transition probabilities are obtained from the distribution of $x(n)$: for all integers i let

$$b_i \triangleq \Pr[|x(n)| \leqq C\gamma^i] \tag{6}$$

and

$$a_i \triangleq 1 - b_i.$$

The "b" is a mnemonic for backward probabilities since it is associated with a transition backwards from the generic state i to $(i - k)$. The diagram in Fig. 2 represents the Markov chain. Denoting by $p_i(n)$ the probability that $\omega(n) = i$, we have

$$\boxed{p_i(n+1) = b_{i+k}p_{i+k}(n) + a_{i-l}p_{i-l}(n).} \tag{7}$$

Although the transition probabilities depend on the distribution of $x(n)$, the two following properties of the sequence $\{b_i\}$, on which we

* D. J. Goodman suggested the above structure on the multiplier coefficients with the object of obtaining a discrete Markov process.

† We are tacitly assuming that $\Pr[|x(n)| = 0] \leqq l/(k+l) - \epsilon$, $\epsilon > 0$.

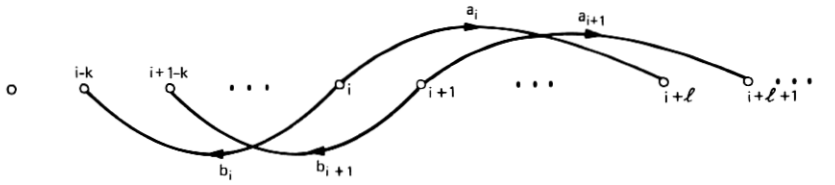

Fig. 2—The Markov chain.

base our results, hold irrespective of the distribution:

$$0 \leqq b_i < b_{i+1} \leqq 1 \quad \text{for all } i, \tag{8}$$

and

$$b_{-1} < \frac{l}{k+l} \leqq b_0. \tag{9}$$

That the strict inequality in (8) holds for all i is a mild restriction on the distribution of $x(n)$; however, certain straightforward modifications may be made to obtain corresponding results when the strict inequality does not hold for all i.

The property of the 0 state to which we alluded earlier may be loosely stated, thus: there is a net drift to the left (right) from states to the right (left) of the 0 state. Formally,

$$\boxed{E[\omega(n+1)\,|\,\omega(n)=i]-i=-(k+l)\left[b_i-\frac{l}{k+l}\right]<0 \text{ if } i>0 \quad >0 \text{ if } i<0.} \tag{10}$$

The above super- and submartingale properties are the basis for the existence of a stochastic Liapunov function (Section 2.2) and the bound obtained in Section 4.2.

Remarks: The random walk in (5) with $k = l = 1$ is also the model for the delta-modulator subject to random, independent, identically distributed inputs. The stationary behavior of the model was treated in an elegant paper by Fine.[9] Gersho[10] has established the stochastic stability of the delta-modulator for a larger class of input processes. Some of our results, particularly those in Section IV on transient response, appear to be new and of some interest in this context.

1.3 The saturating adaptive quantizer

For the algorithm in (1) and, say, Gaussian distributions of the input, there is a small, positive probability of the step size exceeding

any large prespecified level. A model which reflects more accurately the practical algorithm for adapting the step size is one which does not allow the step size to become unbounded. One way of implementing this is to make the step size saturate at some suitably large level, i.e., if $\Delta(n) < |x(n)|$, then

$$\Delta(n+1) = \min [M_2\Delta(n), \bar{L}]; \quad \bar{L} \gg 0; \tag{11}$$

i.e., in the log transformed variables,

$$\omega(n+1) = \min [\omega(n) + l, L]; \quad L \gg 0. \tag{12}$$

The model of this device, which we shall refer to as the saturating adaptive quantizer, is useful not only for the reasons given but also on theoretical grounds since the results obtained for the saturating adaptive quantizer yield, in the limit as $L \to \infty$, corresponding results for the adaptive quantizer. We carry both models with us throughout the paper and at least indicate along the way the main correspondences.

For similar reasons we expect that in practice the step size will also be bounded from below in the obvious manner. This case is not formally dealt with in the text since the main results may be readily inferred from the saturating adaptive quantizer.

For the saturating adaptive quantizer, the following equations govern the evolution of $\{p_i(n) = \Pr [\omega(n) = i]\}$, $i \leqq L$:

$$\begin{aligned} p_i(n+1) &= b_{i+k}p_{i+k}(n) + a_{i-l}p_{i-l}(n) \quad i \leqq L-k \\ p_i(n+1) &= a_{i-l}p_{i-l}(n) \quad L-k+1 \leqq i \leqq L-1 \\ p_L(n+1) &= \sum_{j=L-l}^{L} a_j p_j(n). \end{aligned} \tag{13}$$

The important super- and sub-martingale properties of the random walk, as expressed by the inequalities in eq. (10), apply as well to the saturating adaptive quantizer.

II. THE EXISTENCE AND UNIQUENESS OF THE STATIONARY DISTRIBUTION

We examine in this section questions related to the stochastic stability of the adaptive quantizer. We establish theoretically that certain acute types of erratic operations such as the unboundedness of the evolving random variable, namely, the step size, do not occur. We begin by establishing that the process has the basic properties of a well-behaved process, namely, irreducibility and recurrence. We thereby establish the existence and uniqueness of a finite stationary

distribution. We then proceed to the saturating adaptive quantizer, the more realistic model of the adapting algorithm, which in addition to the above properties, is also aperiodic. Here, the entire state space is a single ergodic class. The main result of this section is obtained from the construction of a stochastic Liapunov function for the process; and the theory of stochastic Liapunov functions is fairly well known.[11,12]

2.1 Irreducibility of the Markov chain

The chain is irreducible if and only if every state communicates with both the neighboring states. This occurs if and only if there exists nonnegative integers m, m', n, n' such that

$$ml - nk = 1 \tag{14a}$$

and

$$m'l - n'k = -1. \tag{14b}$$

It is an elementary fact from number theory that this occurs if and only if k and l are relatively prime, i.e., their greatest common divisor is unity. In fact, Euclid's algorithm yields the unknown quantities in eq. (14).

2.2 Recurrence

Consider the following nonnegative function of the states:

$$V(i) = |i| \qquad i = 0, \pm 1, \cdots. \tag{15}$$

This function is a stochastic Liapunov function[12] if the following holds: if $D(i)$ is defined as follows,

$$E[V\{\omega(n+1)\} \,|\, \omega(n) = i] - V(i) \triangleq D(i), \tag{16}$$

then (i) $D(i)$ is uniformly bounded from above and (ii) $D(i) \leqq -\epsilon < 0$ for all but a finite set of states i. Condition (i) is trivially true for the process. Also, for all $i \geqq k$

$$D(i) = -(k+l)\left(b_i - \frac{l}{k+l}\right) \leqq -(k+l)\left(b_k - \frac{l}{k+l}\right) < 0 \tag{17}$$

and, for all $i \leqq -l$,

$$D(i) = (k+l)\left(b_i - \frac{l}{k+l}\right) \leqq (k+l)\left(b_{-l} - \frac{l}{k+l}\right) < 0. \tag{18}$$

Therefore, condition (ii) is verified, and $V(i)$ is a stochastic Liapunov function for the process.

From Kushner's Theorem 7[12] we have recurrence* and we can infer further, from Theorem 4, that there exists at least one finite invariant measure, i.e., stationary distribution. Also, as we have shown earlier there does not exist two or more disjoint self-contained subsets of the state space; hence, we have from Theorem 5 that there is at most one invariant probability measure. Thus, the existence and uniqueness of a finite stationary distribution for the step size of the adaptive quantizer is established.

2.3 The saturating adaptive quantizer

We will circumvent the technical nuisance† posed by periodicity by proceeding to the saturating adaptive quantizer. In this case the above arguments leading to irreducibility and recurrence are intact. In addition, the end state L has period 1 and, since periodicity is a class concept (i.e., every state in a particular communicating class has the same periodicity), the entire Markov chain is aperiodic. We have, then, $p(n) \to p$ for any $p(0)$ and $p_i > 0$ for all i. Also, the state space is a single ergodic class. Hence, the statistical average of the step sizes approach a limit given by the unique, finite, stationary distribution.

III. SOME PROPERTIES OF THE STATIONARY DISTRIBUTIONS

In this section we investigate in detail properties of the stationary distribution of the step size. In eq. (7) if we set $p_i(n+1) = p_i(n) = p_i$, then the stationary distribution is given by $\{p_i\}$. Thus, the stationary probabilities are the solutions of

$$\boxed{p_i = b_{i+k}p_{i+k} + a_{i-l}p_{i-l}} \tag{19}$$

with, of course, the normalization,

$$\sum_{-\infty}^{\infty} p_i = 1. \tag{20}$$

For the saturating adaptive quantizer, we have from eq. (13) that the basic recursion in (19) holds for all $i \leqq (L-k)$. The remaining

* A Markov chain is recurrent if and only if every state is recurrent; and state i is recurrent if and only if, starting from state i, the probability of returning to state i after some finite length of time is one.

† Feller[13] writes: "The classification into persistent and transient states is fundamental, whereas the classification into periodic and aperiodic states concerns a technical detail."

equations are (20) and the following:

$$p_i = a_{i-l}p_{i-l} \quad L - k + 1 \leqq i \leqq L - 1 \tag{21a}$$

$$p_L = \sum_{L-l}^{L} a_j p_j \tag{21b}$$

and, of course, $p_i = 0$, $i > L$.

3.1 A useful reduction of the equations for the stationary probabilities

To provide some insight into the motivation for the step we undertake here, consider the recursion, analogous to (19), that would arise from a Markov chain with uniform transition probabilities:

$$p_i = bp_{i+k} + ap_{i-l}, \quad a + b = 1. \tag{22}$$

A particular solution of the above recursion is $p_i \equiv c$, a constant. Since, in probability theory, interest is restricted to solutions with bounded sums, one would proceed in the case of (22) by factoring the root at unity from the characteristic polynomial:

$$b\lambda^{k+1} - \lambda^l + a = 0,$$

and thus obtain a new, and reduced, polynomial and an associated recursion. This operation is paralleled for the more general recursion in (19) by the following: from (19),

$$p_i - p_{i-l} = b_{i+k}p_{i+k} - b_{i-l}p_{i-l}.$$

Hence, for all j,

$$\sum_{-\infty}^{j} (p_i - p_{i-l}) = \sum_{-\infty}^{j} (b_{i+k}p_{i+k} - b_{i-l}p_{i-l}) \tag{22a}$$

which reduces to

$$\boxed{\sum_{j+1}^{j+k} b_i p_i = \sum_{j-l+1}^{j} (1 - b_i)p_i.} \tag{23}$$

Remarks:

(*i*) Observe that we are justified in carrying out the operation in (22a) in the case of solutions of (19) for which $\sum_{-\infty}^{j} p_i$ is bounded and which we have established, in Section II, to be the case for the stationary probabilities.

(*ii*) The reduction alluded to earlier refers to the fact that the largest difference in variable indices in (23) is $k + l$, while the largest difference in (19) is $k + l + 1$.

(*iii*) Observe that when $k = l = 1$, (23) gives the solution in closed form: $p_{j+1} = (a_j/b_{j+1})p_j$ and $\sum p_j = 1$. This is a previously known fact; see Feller[14] and Fine.[9] However, neither author gave any indication of the possible generalization to the form in (23).

For the saturating adaptive quantizer, (23) holds for all $j \leqq (L - k)$. Hence, the range over which (23) is valid is such that every state is included in at least one component of the recursion.

3.2 An identity involving the stationary distribution

We use eq. (23) to show that the *stationary* probability of the nth input sample, $x(n)$, falling in the lower slot, $\Pr_s [|x(n)| \leqq \Delta(n)] \equiv l/(k + l)$. The significance of this identity from the point of view of optimal steady-state operation (see Section 1.1) is that by appropriate choice of k and l the above quantity may be matched to the corresponding probability for the optimal nonadaptive step size. This, of course, has the effect of locating the central step size, eq. (3), close to the optimal nonadaptive step size. In the case of independent Gaussian inputs, the above quantity is close to 0.68 and a reasonable approximation is obtained by making $k = 1$ and $l = 2$.

From (23),

$$\sum_{j-l+1}^{j+k} b_i p_i = \sum_{j-l+1}^{j} p_i.$$

Hence,

$$\sum_{j=-\infty}^{\infty} \sum_{i=j-l+1}^{j+k} b_i p_i = \sum_{j=-\infty}^{\infty} \sum_{i=j-l+1}^{j} p_i. \tag{24}$$

The left-hand side equals $(k + l) \sum_{-\infty}^{\infty} b_j p_j$, while the right-hand side equals l. Hence,

$$\boxed{\sum_{-\infty}^{\infty} b_j p_j = \frac{l}{k + l}} \tag{25}$$

Consider what the above equality implies in terms of step size behavior. The stationary probability of the input falling in the lower slot,

$$\Pr_s [|x(n)| \leqq \Delta(n)] = \sum_{i=-\infty}^{\infty} \Pr_s [\Delta = C\gamma^i \quad \text{and} \quad |x| \leqq C\gamma^i]$$

$$= \sum_{i=-\infty}^{\infty} b_i \Pr_s [\Delta = C\gamma^i] \tag{26}$$

from the independence of $\{x(n)\}$. Hence, from (25),

$$\Pr_s [|x(n)| \leqq \Delta(n)] = \frac{l}{k+l}. \tag{27}$$

Immediately on substituting $M_1 = \gamma^{-k}$ and $M_2 = \gamma^l$ we have an identity with a rather appealing and natural interpretation*:

$$M_1^{\rho_1} M_2^{\rho_2} = 1 \tag{28}$$

where ρ_1 and ρ_2 are respectively the two stationary probabilities of the input falling in the lower and upper slots.

For the saturating adaptive quantizer, it can be shown that

$$\sum_{i \leqq L} b_i p_i < \frac{l}{k+l}. \tag{29}$$

However, the quantity $[(l/k + l) - \sum b_i p_i]$ depends only on $(k + l)$ terms involving the end probabilities $p_L, \cdots, p_{L-k-l}$ and it goes to zero with these probabilities. Now we will prove in Section 3.3 certain results which indicate that these probabilities are relatively small if L is large.

3.3 Geometric bounds on the stationary probabilities

In this section we prove a fundamental property of the stationary distribution of the step size which holds for *all* values of γ. We obtain sharp bounds on almost all of the stationary probabilities—the bounds apply as well to the saturating adaptive quantizer—which show that the stationary probability of the random walk being in a particular state falls off at least geometrically with the distance of that state from the 0 state. The actual bounds obtained are substantially stronger and they indicate that a localization property on the stationary distribution is inherent for the random walk. As discussed in Section 1.1 this localization property is important in understanding the basis for the satisfactory behavior of the adaptive quantizer.

We obtain the following point-wise bound: for every $i > 0$ we give positive constants $r > 1$ and c such that for all $j \geqq i$,

$$p_j \leqq c \left(\frac{1}{r}\right)^{j-i}. \tag{30}$$

The quantities r and c depend on i. The quantity r which we call the

---

* D. J. Goodman first conjectured the existence of (28) in the context of the adaptive quantizer. Earlier, N. S. Jayant[5] made a related conjecture in connection with an adaptive delta-modulator.

local steepness factor is a monotonic increasing function of i for nonnegative i. Of course, a corresponding result holds for $i < 0$ and all $j \leqq i$,

Let $\mathbf{P}_i$ denote the $(k + l - 1)$-dimensional column vector* with the following components

$$\mathbf{P}_i \triangleq [p_i, p_{i+1}, \cdots, p_{i+k+l-2}]^t. \tag{31}$$

Then, from (23), we obtain $(k + l - 1) \times (k + l - 1)$ transition matrices $\mathbf{A}_i$, where

$$\mathbf{P}_{i+1} \triangleq \mathbf{A}_i \mathbf{P}_i. \tag{32}$$

The leading $(k + l - 2)$ components of $\mathbf{P}_{i+1}$ are obtained from $\mathbf{P}_i$ by merely shift operations. The nontrivial information in $\mathbf{A}_i$ is in the last row which is obtained from (23); clearly, $\mathbf{A}_i$ depends on i.

We will show that there exist a constant weight vector $\boldsymbol{\lambda}$, every element of $\boldsymbol{\lambda}$ being positive, and a constant $r > 1$ depending only on $\mathbf{A}_i$, such that for all $j \geqq i$

$$\boldsymbol{\lambda}^t \mathbf{A}_j^{-1} \geqq r\boldsymbol{\lambda}^t \tag{33}$$

in the sense that every element of the left vector is not less than the corresponding element of the right vector. Since $\mathbf{P}_{j+1}$ is a vector with nonnegative elements, we have

$$r\boldsymbol{\lambda}^t \mathbf{P}_{j+1} \leqq \boldsymbol{\lambda}^t \mathbf{A}_j^{-1} \mathbf{P}_{j+1} = \boldsymbol{\lambda}^t \mathbf{P}_j. \tag{34}$$

Hence,

$$\boldsymbol{\lambda}^t \mathbf{P}_j \leqq \left(\frac{1}{r}\right)^{j-i} (\boldsymbol{\lambda}^t \mathbf{P}_i) \qquad j \geqq i. \tag{35}$$

Remarks: Equation (35) is a strong result if $\boldsymbol{\lambda}^t \mathbf{P}_j$ is viewed as a norm of the vector $\mathbf{P}_j$ of the L_1-type: $|\mathbf{x}| = \sum \lambda_i |x_i|$, which is a valid interpretation since the latter reduces to $\boldsymbol{\lambda}^t \mathbf{x}$ whenever every element of $\mathbf{x}$ is nonnegative. By standard methods we can obtain upper bounds for $\mathbf{P}_j$ in norms other than the one used in (35). In particular, (30) follows trivially.

It is necessary now to discuss the structure of the matrix $\mathbf{A}_i^{-1}$. Directly from (23) we obtain the first row:†

$$\Big[\overbrace{-\frac{a_{i+1}}{a_i}, -\frac{a_{i+2}}{a_i}, \cdots, \frac{-a_{i+l-1}}{a_i}}^{(l-1) \text{ terms}}, \overbrace{\frac{b_{i+l}}{a_i}, \frac{b_{i+l+1}}{a_i}, \cdots, \frac{b_{i+l+k-1}}{a_i}}^{k \text{ terms}}\Big].$$

* The superscript t denotes the transpose.

† Observe that neither $\mathbf{A}_i$ nor $\mathbf{A}_i^{-1}$ is a stochastic matrix (nonnegative elements, columns sum to unity).

The remaining rows of $\mathbf{A}_i^{-1}$ reflect shift operations: for $m = 2, 3, \cdots, (k + l - 1)$,

$$(\mathbf{A}_i^{-1})_{mn} = 0 \quad \text{if} \quad n \neq (m - 1)$$
$$= 1 \quad \text{if} \quad n = (m - 1).$$

Before proceeding to prove (33) we need the following lemma. This lemma concerns the matrix $\tilde{\mathbf{A}}_i^{-1}$ which is obtained from $\mathbf{A}_i^{-1}$ by merely replacing the first $(l - 1)$ elements of the first row by -1.

Lemma 1: For every $i \geqq 0$

(*i*) $\tilde{\mathbf{A}}_i^{-1}$ has a unique positive real eigenvalue r, say. Furthermore, $r > 1$.

(*ii*) Every element of the corresponding left eigenvector $\boldsymbol{\lambda}$ is of the same sign and nonzero; hence, $\boldsymbol{\lambda}$ may be taken to be a positive vector.

(*iii*) r, which depends on i, is monotonic, strictly increasing with i.

We give the proof of Lemma 1 in Appendix A.

We need one further observation to prove (33) with the help of the lemma. For $j \geqq i$,

$$\boldsymbol{\lambda}^t \mathbf{A}_j^{-1} = \boldsymbol{\lambda}^t(\mathbf{A}_j^{-1} - \tilde{\mathbf{A}}_i^{-1}) + \boldsymbol{\lambda}^t \tilde{\mathbf{A}}_i^{-1}$$
$$= \boldsymbol{\lambda}^t(\mathbf{A}_j^{-1} - \tilde{\mathbf{A}}_i^{-1}) + r\boldsymbol{\lambda}^t.$$

The bound in (33) follows if $\boldsymbol{\lambda}^t(\mathbf{A}_j^{-1} - \tilde{\mathbf{A}}_i^{-1}) \geqq 0$. Since $\boldsymbol{\lambda}$ is a positive vector it is sufficient to show that the elements of the matrix $(\mathbf{A}_j^{-1} - \tilde{\mathbf{A}}_i^{-1})$ are nonnegative. The only nonzero elements of the matrix $(\mathbf{A}_j^{-1} - \tilde{\mathbf{A}}_i^{-1})$ are in the first row. That every term of the first row is nonnegative is implied by the following: for $s \geqq 1$

$$1 - \frac{a_{j+s}}{a_j} \geqq 0 \tag{36}$$

and

$$\frac{b_{j+s}}{a_j} - \frac{b_{i+s}}{a_i} \geqq 0. \tag{37}$$

This concludes the proof of (33) and, hence, of (35).

Remarks:

(*i*) The reader may now appreciate the reason for replacing some of the elements of $\mathbf{A}_i^{-1}$ by -1 to form $\tilde{\mathbf{A}}_i^{-1}$: a_{j+s}/a_j although bounded by 1 can come arbitrarily close to 1.

The reader is also due an explanation for our having worked with $\mathbf{A}_j^{-1}$ after defining the natural transformation $\mathbf{A}_j$, especially since (34) may be put in the form $\lambda^t[\mathbf{I} - r\mathbf{A}_j]\mathbf{P}_j \geqq 0$. The reason is that r and λ, depending only on i, do not exist such that for $j \geqq i$, $\lambda^t[\mathbf{I} - r\mathbf{A}_j] \geqq \mathbf{0}$, although, as we have shown, λ and r do exist such that $\lambda^t[\mathbf{I} - r\mathbf{A}_j]\mathbf{A}_j^{-1} \geqq \mathbf{0}$. In working this step the assumption of $\mathbf{P}_{j+1} \geqq \mathbf{0}$, rather than $\mathbf{P}_j \geqq \mathbf{0}$, appears to be critical.

(*ii*) The interesting quantity $r = r(i)$ may reasonably be called the local steepness factor, since for $i \geqq 0$ it is a local measure of the rapidness with which the stationary distribution falls off. From statement (*iii*) of the lemma we have the fact that the distribution tends to get steeper with increasing distances from the natural center of the distribution, the 0 state.

(*iii*) The theoretical interest in the inequality in (35) results from the fact that we cannot expect to obtain a significantly better value than r for the geometric factor in geometrical bounds on p_j for all $j \geqq i$. The reason for this is that by making b_{j+1} very close to b_j over a fairly large set of j's, it is possible to make the solution of (23) close to the stationary probabilities of a random walk with uniform transition probabilities, which in turn may be obtained in terms of r as the unique positive real root of the characteristic polynomial $C(\mu)$ given in eq. (56), Appendix A.

(*iv*) From symmetry we expect results similar to (35) to hold for $i < 0$. Perhaps the simplest way to show this is by means of the following transformations which have the effect of making the direction of decreasing i the forward direction. Let

$$p'_{-i} = p_i, \quad b'_{-i} = a_i, \quad a'_{-i} = b_i.$$

The basic recursion (23), stated in terms of the new variables, is

$$\sum_{j+1}^{j+l} b'_i p'_i = \sum_{j-k+1}^{j} (1 - b'_i)p'_i.$$

Now $\{b'_i\}$ is a monotonic, increasing sequence with i and $i > 0 \Rightarrow lb'_i > ka'_i$. (Observe the interchange of l and k, i.e., $l' = k$ and $k' = l$.) This transformation makes the transfer of results holding for $i > 0$ to $i < 0$ fairly straightforward.

(*v*) In considering the application of (35) to the saturating adaptive quantizer we note that the basic recursion (23) holds over the entire range of states, i.e., (23) holds for all $j \leqq L - k$. Hence, (35) holds for $L - (l + k) + 2 \geqq j \geqq i \geqq 0$. This observation is the basis for a

statement made earlier in Section 3.2, namely, we expect the tail probabilities of the stationary distribution of the step size for the saturating adaptive quantizer to be small.

From (35) we obtain a rather simple point-wise bound on the stationary probabilities. Let λ_m denote the largest element of the vector $\boldsymbol{\lambda}$. Clearly,*

$$\boldsymbol{\lambda}^t \mathbf{P}_i \leqq \lambda_m \mathbf{1}^t \mathbf{P}_i,$$

and, hence, from (35), for all $j \geqq i \geqq 0$

$$\lambda_m p_{j+m-1} \leqq \boldsymbol{\lambda}^t \mathbf{P}_j \leqq \left(\frac{1}{r}\right)^{j-i} \boldsymbol{\lambda}^t \mathbf{P}_i \leqq \left(\frac{1}{r}\right)^{j-i} \lambda_m (\mathbf{1}^t \mathbf{P}_i),$$

i.e.,

$$p_{j+m-1} \leqq \left(\frac{1}{r}\right)^{j-i} (\mathbf{1}^t \mathbf{P}_i).$$

Hence,

$$\boxed{p_{j+k+l-2} \leqq \left(\frac{1}{r}\right)^{j-i} (\mathbf{1}^t \mathbf{P}_i) \leqq \left(\frac{1}{r}\right)^{j-i}, \quad j \geqq i \geqq 0,} \tag{38}$$

where $r = r(i)$.

3.4 Lower bounds on the steepness factors, r(i)

We have associated with every state i a local steepness factor $r(i)$. Here we go back to the definition of $r(i)$ as being the unique positive root of the polynomial $C(\mu)$, eq. (56), to obtain the following bound which has the advantage of being explicit.

$$\boxed{\left[\frac{kb_i}{la_i}\right]^{1/(k+l-1)} \triangleq \rho(i) \leqq r(i), \qquad i \geqq 0.} \tag{39}$$

Observe that $\rho(i) > 1$ for all $i > 0$ and itself forms a monotonic increasing sequence with i. To prove (39) it is enough to show that $C[(kb_i/la_i)] \leqq 0$. The proof is straightforward but tedious and we omit it.

3.5 The effect of γ on the stationary distribution

We show here that the mass of the stationary distribution of the step size can be localized about the central step size to an arbitrary extent by making γ sufficiently close to unity. To do this, we first put

* The column vector with every element equal to unity is denoted by **1**.

together from the results of the preceding sections a rather explicit bound on the stationary probability of the step size exceeding a particular value for a given γ, i.e., $\Pr_s [\Delta \geqq C\gamma^i]$. This bound is in a form which allows direct comparison with the corresponding probability arising from the choice of $\gamma' = \sqrt{\gamma}$. By successively taking γ to be the square root of the preceding value, the bound on the probability can be made as small as desired. As before, we shall restrict our attention to step sizes which exceed the central step size, i.e., $i > 0$ since a parallel argument holds for $i < 0$.

For $i > 0$ and $r = r(i)$, we have from (35) that

$$(\Sigma\lambda_i) \sum_{j=i+k+l-2}^{\infty} p_j \leqq \sum_{j=i}^{\infty} \mathfrak{A}^t\mathbf{P}_j \leqq \mathfrak{A}^t\mathbf{P}_i \sum_{j=0}^{\infty} \left(\frac{1}{r}\right)^j = \mathfrak{A}^t\mathbf{P}_i \frac{r}{r-1}. \quad (40)$$

Now, as in (39),

$$r \geqq \rho(i) = \left(\frac{kb_i}{la_i}\right)^{1/(k+l-1)}$$

and

$$\frac{\mathfrak{A}^t\mathbf{P}_i}{\Sigma\lambda_i} \leqq \max [p_i, \cdots, p_{i+k+l-2}].$$

Since

$$\Pr_s [\Delta \geqq C\gamma^{i+k+l-2}] = \sum_{j=i+k+l-2}^{\infty} p_j,$$

we have, from (40),

$$\boxed{\Pr_s [\Delta \geqq C\gamma^{i+k+l-2}] \leqq \frac{\rho(i)}{\rho(i)-1} \max [p_i, \cdots, p_{i+k+l-2}].} \quad (41)$$

Finally, from (38), for $i \geqq k + l - 1$,

$$\boxed{\max [p_i, \cdots, p_{i+k+l-2}] \leqq \left(\frac{1}{\rho(1)}\right)^{i-k-l+1}.} \quad (42)$$

Equations (41) and (42) give the bound for the mass of the distribution to the right of a particular state, which we shall now compare with a similar bound that holds for $\gamma' = \sqrt{\gamma}$. The prime superscript will be used on symbols to denote the functional dependence of the associated quantities on γ'. In establishing the reference (central) step size [see eq. (3)], minor differences exist depending on whether

(i) $\Pr [|x(n)| \leqq \gamma^{i-1}] < \dfrac{l}{k+l} \leqq \Pr [|x(n)| \leqq \gamma^{i-1/2}]$

or

$$(ii) \qquad \Pr\left[|x(n)| \leqq \gamma^{i-1/2}\right] < \frac{l}{k+l} \leqq \Pr\left[|x(n)| \leqq \gamma^{i}\right].$$

We consider only (ii), in which case: $\omega'(n) = 2i \Leftrightarrow \omega(n) = i$ and $b'_{2i} = b_i$ for all $i \geqq 0$.

Repeating the arguments leading to (41) and (42) we have

$$\Pr_s\left[\Delta \geqq C\sqrt{\gamma}^{2i+(k+l-2)}\right] \leqq \frac{\rho'(2i)}{\rho'(2i)-1} \max\left[p'_{2i}, \cdots, p'_{2i+k+l-2}\right] \tag{43}$$

and

$$\max\left[p'_{2i}, \cdots, p'_{2i+k+l-2}\right] \leqq \left[\frac{1}{\rho'(2)}\right]^{2i-k-l}. \tag{44}$$

Since $\rho'(2i) = \rho(i)$, we have

$$\Pr_s\left[\Delta \geqq C\sqrt{\gamma}^{2i+(k+l-2)}\right] \leqq \frac{\rho(i)}{\rho(i)-1}\left[\frac{1}{\rho(1)}\right]^{i-k-l+1}\left[\frac{1}{\rho(1)}\right]^{i-1}. \tag{45}$$

Comparison with (41) and (42) completes the demonstration.

IV. TRANSIENT RESPONSE

The preceding section discusses various aspects of the stationary distribution of the step size which effectively describes the steady-state behavior of the device. However, as stated before in Section I, the steady-state response is only of partial interest since the adaptability of the device is tied to quickness of response in the following situations:

- (i) Start up—we are forced to consider situations in which the initial step size is fairly arbitrary.
- (ii) Changes in the scaling of the input distribution—the scenario here is that the device has adapted to a particular intensity level (scaling) of the input distribution when a jump occurs to a new intensity level.

In common with both situations, we have an initial step size and a waiting time for the step size to adapt to the desired step size. Recall that with k and l appropriately chosen, the desirable step size is the central step size, which corresponds to the 0 state in the random walk, eq. (5). This aspect of the behavior of the device is also related to the rate at which the evolving step size distribution approaches the stationary distribution.

The main contribution of this section is the development of formulas for the efficient computation of the mean time required for the step

size to first reach the central step size for various values of the initial step size. The designer can use the information generated by the methods given here in the following manner. Assuming that the designer has some understanding of the rate of variation of the intensity level of the input distribution, he is in a position to determine the smallest value of γ for which the adaptation algorithm adequately tracks the input process. The parameter γ has to be made sufficiently large for the mean waiting time (time, of course, is used synonymously with number of transitions) for adaptation to be small compared to the changes in the location of the desired step size arising from changes in the intensity level.

4.1 The mean time for first passage to the origin

We will consider the random walk, eqs. (5) and (12), for the saturating adaptive quantizer since in the limit, as L becomes large, the functionals obtained for this model yield corresponding quantities for the adaptive quantizer. Also, we shall consider only the case of the initial state $\omega(0) > 0$ since the results obtained can be transferred to the case of negative initial states in a fairly obvious manner (see Remark (*iv*) of Section 3.3).

Let the initial state $\omega(0) = i > 0$ and let M_i denote the mean time required for the first occurrence of the event $\omega(n) \leqq 0$. We observe that for all values of L, not necessarily finite, the time to first passage is finite with probability 1 as a consequence of the properties of recurrence and irreducibility established earlier in Section II. If the first transition results in a decrease of the step size, the process continues as if the initial state has been $(i - k)$. The conditional expectation of the first passage time, therefore, is $M_{i-k} + 1$. From this argument we deduce that the mean first passage time satisfies the recursion*

$$M_i = b_i(M_{i-k} + 1) + a_i(M_{i+l} + 1) \quad \text{for} \quad (k + 1) \leqq i \leqq (L - l). \tag{46}$$

The relation in (46) may be used to generate the entire sequence $\{M_i\}$ provided the initial conditions are known. Now, by the same argument that led to (46), we have that (46) holds for $1 \leqq i \leqq k$ with

$$M_{1-k} = M_{2-k} = \cdots = M_0 = 0.$$

---

* There is some similarity between (46) and the equations arising in gambler's ruin problems[15] and sequential analysis,[16] in which generally $k = l = 1$ and the transition probabilities are not variable.

The remaining l boundary conditions, namely,

$$M_1, M_2, \cdots, M_l$$

are hard to obtain and it is necessary to look more deeply into the dynamics of the process to obtain these quantities.

For every sampling instant we define the L-dimensional vector $\mathbf{z}(n)$ with components $z_j(n)$, $1 \leqq j \leqq L$, where

$$z_j(n) \triangleq \Pr[\omega(n) = j \quad \text{and} \quad \omega(s) \geqq 1 \quad \text{for all } s \leqq n]. \tag{47}$$

These vectors, $\mathbf{z}(n)$, evolve with time according to

$$\mathbf{z}(n+1) = \mathbf{D}\mathbf{z}(n), \quad n \geqq 0. \tag{48}$$

These equations are given in Appendix B. Here we reproduce the structure of the $L \times L$ matrix $\mathbf{D}$:

$$\mathbf{D} = \begin{array}{l} \overbrace{}^{k} \\ \left.\begin{matrix} 0 & \cdots & 0 \\ \vdots & & \\ 0 & & \\ a_1 & & \end{matrix}\right\} l \end{array} \begin{bmatrix} b_{k+1} & & & & & \\ b_{k+2} & \ddots & & & & \\ & \ddots & \ddots & & & \\ & & \ddots & \ddots & & \\ & & & \ddots & \ddots & \\ & & & & \ddots & b_L \\ & & & & & 0 \\ & & & & & \vdots \\ & & & & & 0 \\ & & & a_{L-l} & a_{L-l+1} \cdots & a_L \end{bmatrix} \begin{matrix} \\ \\ \\ a_2 \\ \\ \\ \\ \\ \\ \\ \end{matrix}.$$

Putting together various properties of the matrix $\mathbf{D}$ and the random walk, we obtain, in Appendix B, the following result: for $i \geqq 1$

$$\boxed{\begin{aligned} &M_i = \sum_{j \geqq 1} x_j^{(i)}, \\ &\text{where} \quad [\mathbf{I} - \mathbf{D}]\mathbf{x}^{(i)} = \mathbf{e}_i \end{aligned}} \tag{49}$$

and the elements of the vector $\mathbf{e}_i$ are zero everywhere except at the ith location where it is unity. In Appendix B it is shown that $[\mathbf{I} - \mathbf{D}]$ is nonsingular. We observe parenthetically the virtue of the recursion given in (46) in that it allows us to generate rather easily all the M_i's once the l inversions necessary to evaluate $M_1, \cdots, M_l$ are carried out.

The matrix inversion in (49) may be viewed as a mixed boundary value problem with the first l and the final k equations providing the boundary conditions. The bulk of the elements of the vector $\mathbf{x}^{(i)}$

satisfy a recursion that was encountered previously in Section III:

$$x_j^{(i)} = b_{j+k}x_{j+k}^{(i)} + a_{j-l}x_{j-l}^{(i)}. \tag{50}$$

Furthermore, we show in Appendix B that the elements $x_j^{(i)}$ are all nonnegative. Hence, we are in a position to usefully apply, even for infinite L, the techniques and results of Section III.

First, we carry out the reduction of the equations as stated in Section 3.1 where the motivation for this step is discussed. We obtain

$$\sum_{j=r+1}^{r+k} b_j x_j = \sum_{j=r-l+1}^{r} (1 - b_j)x_j, \quad l \leqq r \leqq (L - k). \tag{51}$$

The superscripts on the x's have not been used since (51) holds for all $\mathbf{x}^{(i)}$, $1 \leqq i \leqq l$.

One benefit of the above form is that it involves one less variable than the original recursion (50). In the important case of $k = 1$ and $l \geqq 1$, this reduction is sufficient to transform the original mixed boundary value problem (49) to an initial value problem, i.e., the solution to the matrix inversion problem (49) satisfies a recursion with specified initial conditions. Exact computation in this case becomes quite trivial. The details of this solution are given in Appendix C. Apart from its independent interest, this result is of particular interest in the adaptive quantizer when the distribution of the input sequence is Gaussian. As discussed previously, it is desirable to have in this case $l/(k + l) = 0.68$, and $k = 1$ and $l = 2$ will suffice.

Another property of the solutions $x^{(i)}$ of (49) which holds for all L is that with increasing j, $x_j^{(i)}$ decreases at least geometrically. This conclusion may be drawn from the bounds obtained in Section 3.3, eqs. (35) and (38). From the point of view of numerical inversion of $[\mathbf{I} - \mathbf{D}]$ for large L, this is a critical property in that it is a necessary condition for most numerical techniques. The reader is referred to Richtmyer and Morton[17] for one such technique that we have used successfully and found to be efficient in that it effectively exploits the band structure of the matrix $[\mathbf{I} - \mathbf{D}]$.

Finally, we remark that while we have dealt exclusively with first passage across the 0 state it is clear that generalizations to first crossings across states other than the 0 state is straightforward.

4.2 Bound on the mean first passage time

Two formulas, eqs. (46) and (49), have been given for computing the mean time required for the step size to adapt from an arbitrary

initial value to the desired, and also central, step size. However, by examining these formulas it is not easy to gain insights into the rate at which this adaptation time grows with the distance separating the two states and its dependence on γ. Here, by probabilistic reasoning, we obtain an explicit upper bound on this time and this bound does provide some insight. As we have done before, we consider here only the case of positive initial states, i.e., $\omega(0) > 0$. Let M_{ij}, $0 \leqq i < j$, denote the mean first passage time under the following conditions: the initial state $\omega(0) = j$ and first crossing occurs after τ transitions if $\omega(\tau) \leqq i$ and $\omega(n) > i$ for all $n < \tau$; then $M_{ij} = E(\tau)$. In this notation the quantity M_j defined in Section 4.1 is equivalent to M_{0j}.

In Section 1.2, eq. (10), it is given that

$$E[\omega(n+1) \mid \omega(n) = i] - i = -(k+l)\left[b_i - \frac{l}{k+l}\right]. \quad (52)$$

Denote the quantity on the right by $-S_i$ and observe that for $i > 0$, $S_{i+1} > S_i > 0$; hence, the supermartingale property. [For the saturating adaptive quantizer, the supermartingale property holds even more strongly, i.e., for $i > 0$, (52) holds with the equality replaced by $\leqq$.] In fact, the supermartingale property holds for the transformed process: $\omega'(n) = \omega(n) + nS_{i+1}$. i.e.,

$$E[\omega'(n+1) \mid \omega'(n)] \leqq \omega'(n) \quad (53)$$

for all $\omega'(n) \geqq (i+1) + nS_{i+1}$. For the crossing problem, (53) holds for all $(n+1) \leqq \tau$, the crossing time. We can now apply a theorem due to Doob[18] on optional stopping on supermartingales. In this case, the theorem states that

$$E[\omega'(\tau)] \leqq E[\omega'(0)]. \quad (54)$$

Since

$$(i+1-k) + S_{i+1}E(\tau) \leqq E[\omega'(\tau)] \leqq E[\omega'(0)] = j,$$

we obtain

$$\boxed{M_{ij} = E(\tau) \leqq \frac{1}{S_{i+1}}[(j-i) + (k-1)].} \quad (55)$$

We gain some insight on the role of γ in determining the transient response of the device by observing the dependence of the above bound on γ. Suppose we are interested in M_{0j}, the waiting time for the initial

step size $\Delta(0) = C\gamma^j$ to reach the central step size C. Consider the effects of making $\gamma' = \sqrt{\gamma}$ on this waiting time (the multiplier coefficients of the device are therefore $\sqrt{\gamma}^{-k}$ and $\sqrt{\gamma}^{l}$). We let the prime superscript on symbols indicate a functional dependence on γ'. In establishing the new central step size [see eq. (3)], minor differences exist depending on whether

$$(i) \quad \Pr\left[|x(n)| \leqq \gamma^{i-1}\right] < \frac{l}{k+l} \leqq \Pr\left[|x(n)| \leqq \gamma^{i-\frac{1}{2}}\right]$$

or

$$(ii) \quad \Pr\left[|x(n)| \leqq \gamma^{i-\frac{1}{2}}\right] < \frac{l}{k+l} \leqq \Pr\left[|x(n)| \leqq \gamma^{i}\right].$$

We consider only (ii), in which case the central step sizes are identical: $\omega'(n) = 2i \Leftrightarrow \omega(n) = i$ and $b'_{2i} = b_i$ for all $i \geqq 0$. The waiting time

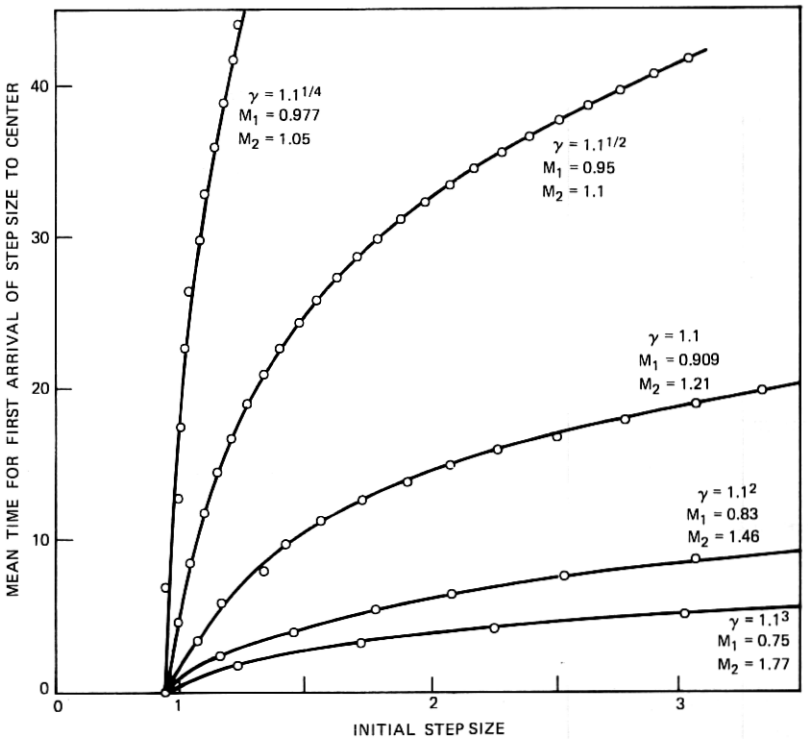

Fig. 3—Transient response of the adaptive quantizer.

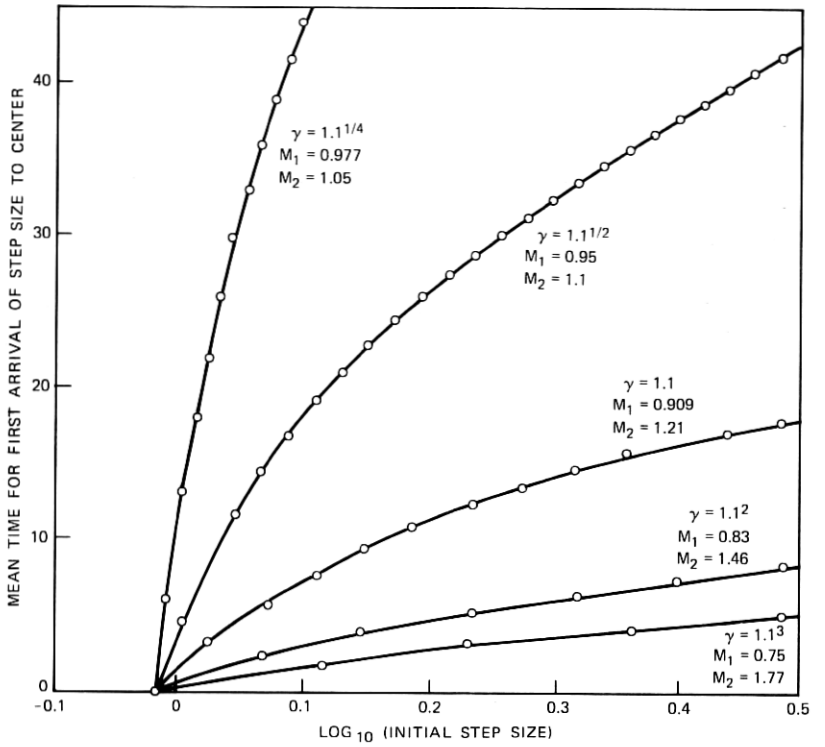

Fig. 4—Transient response of the adaptive quantizer.

for the step size to adapt from identical initial step size $C\gamma^j$ to final step size C is $M'_{0,2j}$. From (55),

$$M'_{0,2j} \leqq \frac{1}{S'_1}[2j - (k-1)].$$

Now, $S_0 \leqq S'_1 \leqq S_1$; hence, making $\gamma' = \sqrt{\gamma}$ and keeping k and l unchanged has the effect of making the bound on the waiting time at least twice as large for $j \gg k$. This is a conclusion which is plausible in the light of the linear form of the bound (55) since the effect of making $\gamma' = \sqrt{\gamma}$ is to introduce twice as many transitions between the initial and final step sizes.

4.3 Computational results

We present here a sampling of our computational results. It is assumed that for every n, $x(n)$ is normally distributed with unit

variance. The optimal step size $\hat{\Delta}$ in this case has the property that $\Pr\{|x(n)| \leqq \hat{\Delta}\} = 0.68$. To center the stationary distribution of the step size close to the optimal step size, we choose $k = 1$ and $l = 2$.

Figure 3 plots the mean time for first passage to the optimal step size vs. initial step size, and the initial step sizes chosen for this figure exceed the optimal step size. Various values of $\gamma(M_1 = \gamma^{-k}, M_2 = \gamma^{l})$ were used. Figure 4 provides the same information except that the horizontal axis corresponds to $\log_{10} \Delta(0)$, rather than $\Delta(0)$ as in Fig. 3. The mean first passage times M_1 and M_2 were obtained by the method outlined in Appendix C, and M_i, $i \geqq 3$ were generated by using the recursion in (46). To give some idea of the rate of convergence for $x_j^{(1)}$, eqs. (70) and (71), we tabulate some values of $x_j^{(1)}$ for the case of

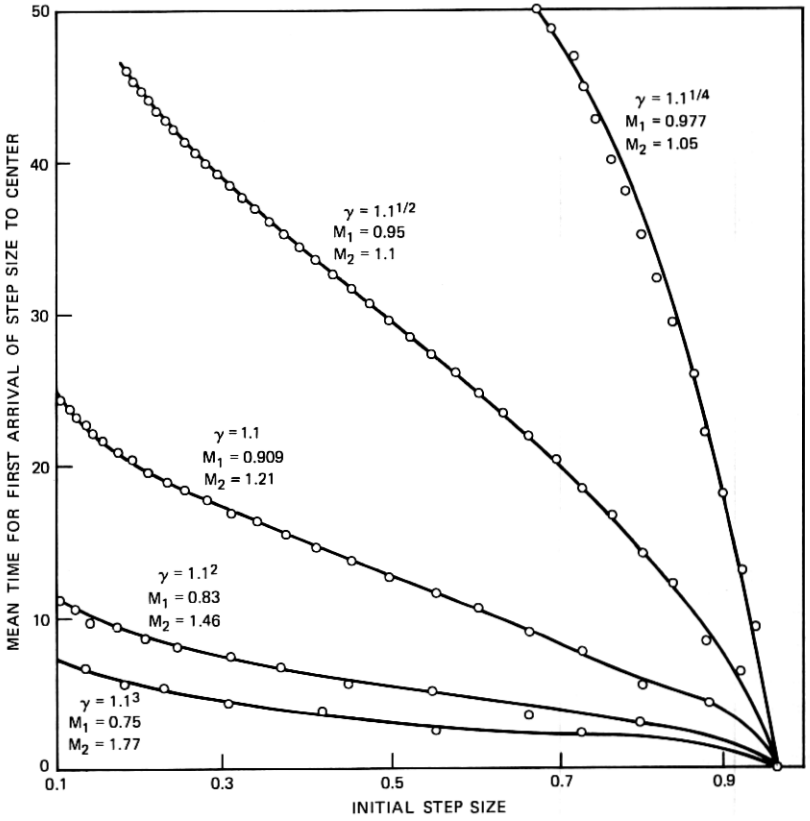

Fig. 5—Transient response of the adaptive quantizer.

$\gamma = 1.1$:

j:	1	2	3	4	5	6	7	8	9	10
$x_j^{(0)}$:	1.4	0.53	0.66	0.31	0.20	0.08	0.03	0.92×10^{-2}	0.24×10^{-2}	0.41×10^{-3}

j:	11	12	13	14	15	16
$x_j^{(0)}$:	0.59×10^{-4}	0.53×10^{-5}	0.35×10^{-6}	0.13×10^{-7}	0.30×10^{-9}	0.31×10^{-11}

Figure 5 is similar to Fig. 3 except that here the initial step sizes are less than the optimal step size. Figure 6 plots the same information with $\log_{10} \Delta(0)$, rather than $\Delta(0)$, on the horizontal axis. The mean first passage time M_1 was obtained by solving (49) by the method given in Ref. 17 and all other first passage times were generated by the recursion in (46).

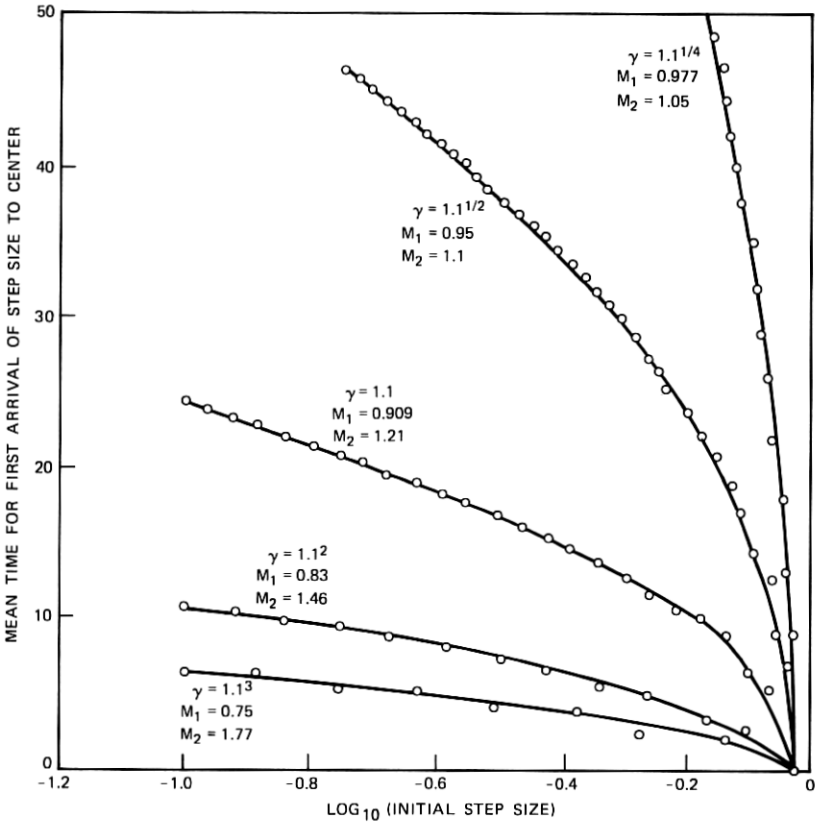

Fig. 6—Transient response of the adaptive quantizer.

V. ACKNOWLEDGMENTS

The criticism and comments of my colleague, Mohan Sondhi, readily and generously given, were always beneficial to me. I benefited greatly from David Goodman's insight while the problem was being formulated. I am grateful to Larry Shepp for the information that Doob's theorem on optional stopping of supermartingales can be used effectively in obtaining the bound on the mean first passage time given in Section 4.2. The benefit of stimulating discussions with N. S. Jayant is acknowledged.

APPENDIX A

Proof of Lemma 1

Proof:

(*i*) $\tilde{\mathbf{A}}_i^{-1}$ being in the form of a companion matrix, the coefficients of the characteristic polynomial of the matrix are the elements of the first row:

$$\begin{aligned} C(\mu) &\triangleq (-1)^{k+l-1} \det [\tilde{\mathbf{A}}_i^{-1} - \mu \mathbf{I}] \\ &= \mu^{k+l-1} + \cdots + \mu^k - [\alpha_1 \mu^{k-1} + \alpha_2 \mu^{k-2} + \cdots + \alpha_k], \end{aligned} \tag{56}$$

where

$$\alpha_1 = \frac{b_{i+l}}{a_i}, \quad \alpha_2 = \frac{b_{i+l+1}}{a_i}, \quad \cdots, \quad \alpha_k = \frac{b_{i+l+k-1}}{a_i}. \tag{57}$$

By Descartes's rule the polynomial $C(\mu)$ has at most one positive real root. Since $C(0) = -\alpha_k < 0$ and $C(\mu) \to \infty$ as $\mu \to \infty$, there exists exactly one positive root. Let r denote this root.

Now $C(1) < 0$ if $la_i < (b_{i+l} + b_{i+l+1} + \cdots + b_{i+l+k-1})$. The latter condition holds for all $i \geqq 0$. Hence, $r > 1$.

(*ii*) The left eigenvector $\boldsymbol{\lambda}$ corresponding to the eigenvalue r satisfies, by definition, $\boldsymbol{\lambda}^t \tilde{\mathbf{A}}_i^{-1} = r\boldsymbol{\lambda}^t$. Examining the component equations we find that

$$\lambda_i = \lambda_1(1 + r + \cdots + r^{i-1}) \quad 1 \leqq i \leqq l. \tag{58}$$

Also,

$$\lambda_{l+k-i} = \frac{\lambda_{l+k-1}}{\alpha_k r^{i-1}}[\alpha_{k-i+1} r^{i-1} + \cdots + \alpha_{k-1} r + \alpha_k] \quad 1 \leqq i \leqq k. \tag{59}$$

Finally, $r\lambda_{l+k-1} = \alpha_k \lambda_1$. Since the α's are positive quantities, the statement is clearly true.

(*ii*) The statement can be verified by inspecting the characteristic polynomial $C(\mu)$ and using the fact that the coefficients $\alpha_1, \cdots, \alpha_k$ each increase with i.

APPENDIX B

Derivation of equations (48) and (49)

The derivation of the equations governing the evolution of the vectors $\mathbf{z}(n)$ defined in eq. (47) proceeds as follows. For convenience, let $X(n)$ denote the event $1 \leqq \omega(\tau) \leqq L$ for all τ, $0 \leqq \tau \leqq n$. Hence, by definition,

$$z_j(n) = \Pr[\omega(n) = j \quad \text{and} \quad X_n] \quad 1 \leqq j \leqq L.$$

Since

$$\begin{aligned} z_j(n) &= \Pr[\omega(n) = j \quad \text{and} \quad X_{n-1}] \\ &= \sum_{i=1}^{L} \Pr[\omega(n) = j \,|\, \omega(n-1) = i,\ X_{n-1}] z_i(n-1) \\ &= \begin{cases} b_{k+j}\, z_{k+j}(n-1), & 1 \leqq j \leqq l, \\ a_{j-l}\, z_{j-l}(n-1) + b_{j+k} z_{j+k}(n-1), & (l+1) \leqq j \leqq (L-k) \\ a_{j-l}\, z_{j-l}(n-1), & (L-k+1) \leqq j \leqq (L-1), \\ \sum_{i=L-l}^{L} a_i z_i(n-1) & j = L. \end{cases} \end{aligned}$$

The above equations define the matrix $\mathbf{D}$ which relates $\mathbf{z}(n)$ to $\mathbf{z}(n-1)$ as in eq. (48).

For the derivation of eq. (49) we proceed as follows. For $i = 1, 2, \cdots, L$, let

$$\begin{aligned} F_i(n+1) &\triangleq \Pr[\text{first passage occurs at } (n+1) \,|\, \omega(0) = i] \\ &= \Pr[\omega(n+1) \leqq 0, \quad X_n \,|\, \omega(0) = i] \\ &= \sum_{j=1}^{k} b_j\, z_j(n) \quad \text{with} \quad \mathbf{z}(0) = \mathbf{e}_i. \end{aligned} \tag{60}$$

The vector $\mathbf{e}_i$ has every element equal to zero except for the ith element which is unity. To express eq. (60) in vector form we let $\mathbf{b} \triangleq [b_1 b_2 \cdots b_k\ 0 \cdots 0]^t$. Then, from (60),

$$F_i(n+1) = \mathbf{b}^t \mathbf{z}(n) \quad \text{with} \quad \mathbf{z}(0) = \mathbf{e}_i.$$

By definition, we have that the mean first passage time conditional on the initial state being i,

$$M_i = \sum_{n \geqq 0} (n+1)F_i(n+1)$$

$$= \mathbf{b}^t \sum_{n \geqq 0} (n+1)\mathbf{z}(n)$$

$$= \mathbf{b}^t \sum_{n \geqq 0} n\mathbf{z}(n) + \mathbf{b}^t \sum_{n \geqq 0} \mathbf{z}(n). \tag{61}$$

Now the second term in the above expression is unity since the probability that passage occurs at finite time is unity. Now consider

$$[\mathbf{I} - \mathbf{D}] \sum_{n \geqq 1} n\mathbf{z}(n) = \sum_{n \geqq 1} n\mathbf{z}(n) - \sum_{n \geqq 1} n\mathbf{z}(n+1)$$

$$= \sum_{n \geqq 0} \mathbf{z}(n) - \mathbf{z}(0). \tag{62}$$

Hence, denoting by $\mathbf{1}$ the column vector with every element equal to unity, we have from (62) that

$$\mathbf{1}^t[\mathbf{I} - \mathbf{D}] \sum_{n \geqq 1} n\mathbf{z}(n) = \mathbf{1}^t \sum_{n \geqq 0} \mathbf{z}(n) - 1 \tag{63}$$

$$= \mathbf{b}^t \sum_{n \geqq 1} n\mathbf{z}(n), \tag{64}$$

since $\mathbf{1}^t\mathbf{z}(0) = 1$ and $\mathbf{b}^t = \mathbf{1}^t[\mathbf{I} - \mathbf{D}]$. It only remains to consider

$$\sum_{n \geqq 0} \mathbf{z}(n) = \left[\sum_{i=0}^{\infty} \mathbf{D}^i \right] \mathbf{z}(0).$$

The above series converges since every eigenvalue of the matrix $\mathbf{D}$ lies strictly within the unit circle in the complex plane. The proof of this follows from an old matrix theorem[19] which states that if the diagonal elements of the columns weakly dominate the sum of the absolute values of the off-diagonal elements with strong dominance holding for at least one column and the matrix is irreducible, then the determinant is nonzero. Applying this theorem to $[\mathbf{D} - \lambda\mathbf{I}]$, $|\lambda| \geqq 1$, we note that the irreducibility of the original Markov chain implies irreducibility of the matrix $[\mathbf{D} - \lambda\mathbf{I}]$ and that the weak column dominance property holds everywhere while the strong column dominance property holds for the first k columns. Hence,

$$\sum_{n \geqq 0} \mathbf{z}(n) = \left[\sum_{i \geqq 0} \mathbf{D}^i \right] \mathbf{z}(0) = [\mathbf{I} - \mathbf{D}]^{-1}\mathbf{z}(0). \tag{65}$$

Putting together the above results we have (49), namely,

$$M_i = \sum_{j \geqq 1} x_j^{(i)} \quad \text{where} \quad [\mathbf{I} - \mathbf{D}]\mathbf{x}^{(i)} = \mathbf{e}_i.$$

Observe that $\mathbf{x}^{(i)} = \Sigma \mathbf{z}(n)$ and, from the definition of $\mathbf{z}(n)$, it follows that every element of $\mathbf{x}^{(i)}$ is nonnegative.

APPENDIX C

Mean first passage times for the case $k = 1$, $l \geqq 1$

We have as our starting point eq. (49), namely,

$$M_i = \sum_j x_j^{(i)}, \tag{66}$$

$$\text{where} \quad [\mathbf{I} - \mathbf{D}]\mathbf{x}^{(i)} = \mathbf{e}_i \tag{67}$$

and we are interested only in $1 \leqq i \leqq l$.

The transformation that was made in Section 3.1 is equivalent to the following: add to each row, r, of $[\mathbf{I} - \mathbf{D}]$ all rows $r + 1, r + 2, \cdots$; and do the same to the vector $\mathbf{e}_i$. This operation makes the matrix $[\mathbf{I} - \mathbf{D}]$ lower triangular, the reason being that with the exception of the first column, the elements of all other columns of $[\mathbf{I} - \mathbf{D}]$ sum to zero. The resulting equations are as follows: the first component equation yields

$$b_1 x_1^{(i)} = 1, \tag{68}$$

and the next $(l - 1)$ equations: $2 \leqq r \leqq l$,

$$-\sum_{j=1}^{r-1} a_j x_j^{(i)} + b_r x_r = \begin{cases} 1 & \text{if} \quad r \leqq i \\ 0 & \text{if} \quad r > i. \end{cases} \tag{69}$$

Finally,

$$\boxed{x_r^{(i)} = \frac{1}{b_r} \sum_{j=r-l}^{r-1} a_j x_j^{(i)} \quad \text{for} \quad r > l.} \tag{70}$$

The boundary conditions to the basic recursion in (70) are in (68) and (69) which are, of course, solvable:

$$\boxed{\begin{aligned} 1 \leqq r \leqq i \quad & x_r^{(i)} = 1 / \prod_{j=1}^{r} b_j \\ (i+1) \leqq r \leqq l \quad & x_r^{(i)} = (x_i^{(i)} - 1) / \prod_{j=i+1}^{r} b_j. \end{aligned}} \tag{71}$$

REFERENCES

1. N. S. Jayant, "Adaptive Quantization with a One-Word Memory," B.S.T.J., *52*, No. 7 (September 1973), pp. 1119–1144.
2. P. Cummiskey, N. S. Jayant, and J. L. Flanagan, "Adaptive Quantization in Differential PCM Coding of Speech," B.S.T.J. *52*, No. 7 (September 1973), pp. 1105–1118.
3. M. R. Winkler, "High Information Delta Modulation," IEEE Int. Conv. Record, part 8, 1963, pp. 260–265.
4. J. E. Abate, "Linear and Adaptive Delta Modulation," Proc. IEEE, March 1967, pp. 298–307.
5. N. S. Jayant, "Adaptive Delta Modulation with a One-Bit Memory," B.S.T.J., *49*, No. 3 (March 1970), pp. 321–342.
6. R. M. Wilkinson, "An Adaptive Pulse Code Modulator for Speech," Proc. Int. Conf. Commun., Montreal, June 1971, pp. 1–11 to 1–15.
7. D. J. Goodman and A. Gersho, "Theory of an Adaptive Quantizer," Proc. of December 1973 IEEE Symp. on Adaptive Processes, Decision and Control.
8. J. Max, "Quantization for Minimum Distortion," Trans. IRE, *IT-6* (March 1960), pp. 7–12.
9. T. Fine, "The Response of a Particular Nonlinear System with Feedback to Each of Two Random Processes," IEEE Trans. Inform. Theory, *IT-14*, No. 2 (March 1968), pp. 255–264.
10. A. Gersho, "Stochastic Stability of Delta Modulation," B.S.T.J., *51*, No. 4 (April 1972), pp. 821–842.
11. R. S. Bucy, "Stability and Positive Super-Martingales," J. Differential Equations, *1*, 1965, pp. 151–155.
12. H. Kushner, *Introduction to Stochastic Control*, New York: Holt, Rinehart and Winston, 1971.
13. W. Feller, *Introduction to Probability Theory and Its Applications*, vol. 1, 3rd ed., New York: John Wiley & Sons, 1950, p. 387.
14. Ref. 13, p. 402.
15. Ref. 13, pp. 348–349.
16. A. Wald, *Sequential Analysis*, New York: John Wiley & Sons, 1947.
17. R. D. Richtmyer and K. W. Morton, *Difference Methods for Initial-Value Problems*, 2nd ed., New York: John Wiley & Sons, 1957, pp. 198–201.
18. J. L. Doob, *Stochastic Processes*, New York: John Wiley & Sons, 1953, pp. 300–301.
19. M. Marden, "Geometry of Polynomials," Mathematical Surveys 3, American Mathematical Society, Providence, R.I., 1966, pp. 140–141.